Scientiae Mathematicae Japonicae Online, **Vol. 10**, (2004), 235–258

DISCRIMINANT ANALYSIS FOR MULTIVARIATE NON-GAUSSIAN LOCALLY STATIONARY PROCESSES

Junichi HIRUKAWA*

Received February 5, 2004; revised March 3, 2004

Abstract. The extension of classical discriminant analysis techniques in multivariate analysis to time series data is a problem of practical interest. Discrimination between different classes of multivariate locally stationary processes, which constitute a class of non-stationary processes, can be characterized by differing covariance or time varying spectral structures. For discrimination between the multivariate non-Gaussian locally stationary processes, Kullback-Leibler discrimination information measure has been developed. In this paper, asymptotic error rates and limiting distributions are given for a generalized time varying spectral disparity measure that includes foregoing criteria as special case. It is well known that the log-likelihood ratio based on observed stretch gives optimal classification. It is shown that the discriminant criterion based on such generalized disparity measure is Gaussian optimal. A non-Gaussian optimal discriminant criterion is also proposed in view of the LAN theorem.

1. Introduction. In multivariate analysis, many methods of discriminant analysis have been investigated in detail (e.g. Anderson, 1984). The extension of classical discriminant analysis techniques in multivariate analysis to time series data is a problem of practical interest. Shumway (1982) gave an extensive review of various discriminant problems in time series. Zhang and Taniguchi (1994) discussed the parametric discriminant problems for non-Gaussian vector linear processes, and showed that discriminant criterion based on an integral functional of periodogram has some good properties, for example, asymptotic normality and non-Gaussian robustness, etc. Zhang and Taniguchi (1995) have shown robustness of Chernoff information measure to peak contamination in spectra of the process concerned. For discrimination between non-Gaussian multivariate time series, Kakizawa et al. (1998) have introduced a disparity measure, which includes the Kullback-Leibler discrimination information and the Chernoff information measure, and gave applications to the problems of classifying earthquakes and mining explosions.

Although the analysis for stationary time series is well established, stationary time series models are not plausible to describe the real world. Dahlhaus (1996a, 1996b, 1996c, 1997) has introduced an important class of non-stationary processes, called locally stationary processes, and established the asymptotic theory of statistical inference. Discrimination between different classes of multivariate time series that can be characterized by differing covariance or time varying structures is important in applications of occurring in the analysis of seismic records and biometrics data. Sakiyama and Taniguchi (2004) investigated the problems of classifying a multivariate non-Gaussian locally stationary process $\{\mathbf{X}_{t,T}\}$ into one of two categories described by two hypotheses $\Pi_i : \mathbf{f}_i(u, \lambda)$, $i = 1, 2$, where $\mathbf{f}_i(u, \lambda)$ are time varying spectral density matrices. They used an approximation of the Gaussian Kullbuck-Leibler information measure as a classification statistic for this problem and showed that this statistic is consistent. The misclassification probabilities were

2000 *Mathematics Subject Classification.* 62M10; 62M15; (62H30).

Key words and phrases. Locally stationary vector process; Discriminant analysis; Chernoff; Kullback-Leibler; Non-Gaussian robust; Peak robustness; Non-Gaussian optimal.

also evaluated under contiguous hypotheses. In this paper, we generalize this measure to non-linear time varying spectral measures which include the Kullbuck-Leibler information and the Cernoff information measure. We also propose a genuine non-Gaussian optimal discrimination criterion based on another approach.
The time series data recoded in real phenomena such as seismic record and financial time series, are often non-stationary and non-Gaussian. To investigate the actual performance of our discrimination criterion to such multivariate non-stationary and non-Gaussian time series data will be increasing importance. However, this problem requires another paper. We will make it as a future work.
This paper is organized as follows. In Section 2 we define the multivariate non-Gaussian locally stationary processes, and introduce a nonparametric time varying spectral density estimator, which is due to Dahlhaus (1996a, 1996b, 1997). Section 3 gives a generalized measure of disparity which includes Kullbuck-Leibler and Chernoff information measure. In Section 4, we derive the limiting distributions and asymptotic error rates of our discriminant statistics. We also discuss conditions for non-Gaussian robustness, and show that our discriminant criterion is Gaussian optimal. Peak robustness of our discrimination criterion is studied, and some numerical examples are given. In Section 5, we propose a genuine non-Gaussian optimal discrimination criterion based on the LAN property. All the proofs of Theorems are given in Section 6.

2. Non parametric spectral estimator of multivariate locally stationary processes. When we deal with nonstationary processes, one of the difficult problems to solve is how to set up an adequate asymptotic theory. To meet this Dahlhaus (1996a, 1996b, 1997) introduced an important class of nonstationary processes and developed the statistical inference. We give the precise definition of multivariate locally stationary processes which is due to Dahlhaus (2000).

Definition 1. *A sequence of multivariate stochastic processes* $\mathbf{X}_{t,T} = (X_{t,T}^{(1)}, \ldots, X_{t,T}^{(m)})'$, $(t = 2 - N/2, \ldots, 1, \ldots, T, \ldots, T + N/2; T, N \geq 1)$ *is called locally stationary with mean vector* $\mathbf{0}$ *and transfer function matrix* $\mathbf{A}^\circ$ *if there exists a representation*

$$\mathbf{X}_{t,T} = \int_{-\pi}^{\pi} \exp(i\lambda t) \mathbf{A}_{t,T}^\circ(\lambda) d\xi(\lambda), \tag{1}$$

where

(i) $\xi(\lambda) = (\xi_1(\lambda), \ldots, \xi_m(\lambda))'$ *is a complex valued stochastic vector process on* $[-\pi, \pi]$ *with* $\overline{\xi_a(\lambda)} = \xi_a(-\lambda)$ *and*

$$cum\{d\xi_{a_1}(\lambda_1), \ldots, d\xi_{a_k}(\lambda_k)\} = \eta(\sum_{j=1}^{k} \lambda_j) \frac{\kappa_{a_1,\ldots,a_k}}{(2\pi)^{k-1}} d\lambda_1 \ldots d\lambda_{k-1}, \tag{2}$$

for $k \geq 2$, $a_1, \ldots, a_k = 1, \ldots, m$, *where* $cum\{\ldots\}$ *denotes the cumulant of* k*-th order, and* $\eta(\lambda) = \sum_{j=-\infty}^{\infty} \delta(\lambda + 2\pi j)$ *is the period* 2π *extension of the Dirac delta function.*

(ii) *There exists a constant* K *and a* 2π*-periodic matrix valued function* $\mathbf{A} : [0,1] \times \mathbb{R} \to \mathbb{C}^{m \times m}$ *with* $\mathbf{A}(u, -\lambda) = \overline{\mathbf{A}(u, \lambda)}$ *and*

$$\sup_{t,\lambda} \left| \mathbf{A}_{t,T}^\circ(\lambda)_{a,b} - \mathbf{A}\left(\frac{t}{T}, \lambda\right)_{a,b} \right| \leq KT^{-1} \tag{3}$$

for all $a, b = 1, \ldots, m$ *and* $T \in \mathbb{N}$. $\mathbf{A}(u, \lambda)$ *is assumed to be continuous in* u.

$\mathbf{f}(u,\lambda) := \mathbf{A}(u,\lambda)\mathbf{\Omega}\mathbf{A}(u,\lambda)^*$ is called the time varying spectral density matrix of the process, where $\mathbf{\Omega} = (\kappa_{a,b})_{a,b=1,\dots,m}$ and $\mathbf{D}^*$ denotes the complex conjugate of matrix $\mathbf{D}$. Write

$$\varepsilon_t := \int_{-\pi}^{\pi} \exp(i\lambda t) d\xi(\lambda), \tag{4}$$

then $E(\varepsilon_t) = \mathbf{0}$, $E(\varepsilon_t \varepsilon_t') = \mathbf{\Omega}$ and $E(\varepsilon_t \varepsilon_s')$ for $t \neq s$ is a zero matrix. We make the following assumption.

Assumption 1. $\mathbf{X}_{t,T}$ *has the MA(∞) representation*

$$\mathbf{X}_{t,T} = \sum_{s=-\infty}^{\infty} \mathbf{a}_{t,T}(s)\varepsilon_{t-s}, \tag{5}$$

that is,

$$\mathbf{A}^{\circ}_{t,T}(\lambda) = \sum_{s=-\infty}^{\infty} \mathbf{a}_{t,T}(s) \exp(-i\lambda s), \tag{6}$$

where the coefficients fulfill

$$\sup_t \sum_{s=-\infty}^{\infty} \left| \left\{ \mathbf{a}_{t,T}(s) - \mathbf{a}_s\left(\frac{t}{T}\right) \right\}_{cd} \right| = O(T^{-1}) \tag{7}$$

for all $c, d = 1, \dots m$, with continuous matrix function $\mathbf{a}_s(u)$. Then, the condition (3) is fulfilled for

$$\mathbf{A}(u,\lambda) = \sum_{s=-\infty}^{\infty} \mathbf{a}_s(u) \exp(-i\lambda s). \tag{8}$$

Furthermore we make the following assumption on the transfer function matrix $\mathbf{A}(u,\lambda)$.

Assumption 2. **(i)** *The transfer function matrix $\mathbf{A}(u,\lambda)$ is 2π-periodic in λ and the periodic extension is twice differentiable in u and λ with uniformly bounded continuous derivatives $\frac{\partial^2}{\partial u^2}\mathbf{A}$, $\frac{\partial^2}{\partial \lambda^2}\mathbf{A}$ and $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial \lambda}\mathbf{A}$.*

(ii) *All the eigenvalues of $\mathbf{f}(u,\lambda)$ are bounded from below and above by some constants $\delta_1, \delta_2 > 0$ uniformly in u and λ.*

As an estimator of $\mathbf{f}(u,\lambda)$, we use the nonparametric estimator of kernel type defined by

$$\widehat{\mathbf{f}}_T(u,\lambda) = \int_{-\pi}^{\pi} W_T(\lambda-\mu)\mathbf{I}_N(u,\mu)d\mu, \tag{9}$$

where $W_T(\omega) = M\sum_{\nu=-\infty}^{\infty} W(M(\omega + 2\pi\nu))$ is the weight function and $M > 0$ depends on T, and $\mathbf{I}_N(u,\lambda)$ is the data tapered periodogram matrix over the segment $\{[uT]-N/2+1, [uT]+N/2\}$ defined as

$$\begin{aligned} \mathbf{I}_N(u,\lambda) = \frac{1}{2\pi H_{2,N}} &\left\{ \sum_{s=1}^{N} h\left(\frac{s}{N}\right) \mathbf{X}_{[uT]-N/2+s,T} \exp\{i\lambda s\} \right\} \\ &\cdot \left\{ \sum_{r=1}^{N} h\left(\frac{r}{N}\right) \mathbf{X}_{[uT]-N/2+r,T} \exp\{i\lambda r\} \right\}^* . \end{aligned} \tag{10}$$

Here $h : [0,1] \to \mathbb{R}$ is a data taper and $H_{2,N} = \sum_{s=1}^{N} h\left(\frac{s}{N}\right)^2$. It should be noted that $\mathbf{I}_N(u,\lambda)$ is not a consistent estimator of the spectral density. To make a consistent estimator of $\mathbf{f}(u,\lambda)$ we have to smooth it over neighbouring frequencies.
Now we impose the following assumptions on $W(\cdot)$ and $h(\cdot)$.

Assumption 3. *The weighted function $W : \mathbb{R} \to [0,\infty]$ satisfies $W(x) = 0$ for $x \notin [-1/2, 1/2]$, and is a continuous and even function satisfying $\int_{-1/2}^{1/2} W(x)dx = 1$ and $\int_{-1/2}^{1/2} x^2 W(x)dx < \infty$.*

Assumption 4. *The data taper $h : \mathbb{R} \to \mathbb{R}$ satisfies (i) $h(x) = 0$ for all $x \notin [0,1]$ and $h(x) = h(1-x)$, (ii) $h(x)$ is continuous on $\mathbb{R}$, twice differentiable at all $x \notin U$ where U is a finite set of $\mathbb{R}$, and $\sup_{x\notin U} |h''(x)| < \infty$. Write*

$$K_t(x) := \left\{\int_0^1 h(x)^2 dx\right\}^{-1} h(x+1/2)^2, \quad x \in [-1/2, 1/2], \tag{11}$$

which plays a role of kernel in the time domain.

Furthermore, we assume

Assumption 5. *$M = M(T)$ and $N = N(T)$, $M \ll N \ll T$ satisfy*

$$\frac{\sqrt{T}}{M^2} = o(1), \ \frac{N^2}{T^{\frac{3}{2}}} = o(1), \ \frac{\sqrt{T}\log N}{N} = o(1). \tag{12}$$

The following lemmas are multivariate version of Theorem 2.2 of Dahlhaus (1996c) and Theorem A.2 of Dahlhaus (1997) (See also Sakiyama and Taniguchi (2003)).

Lemma 1. *Assume that Assumptions 1-5 hold. Then*

(i)

$$\begin{aligned} E(\mathbf{I}_N(u,\lambda)) &= \mathbf{f}(u,\lambda) + \frac{N^2}{2T^2}\int_{-1/2}^{1/2} x^2 K_t(x)^2 dx \frac{\partial^2}{\partial u^2}\mathbf{f}(u,\lambda) \\ &\quad + o\left(\frac{N^2}{T^2}\right) + O\left(\frac{\log N}{N}\right), \end{aligned} \tag{13}$$

(ii)

$$\begin{aligned} E(\widehat{\mathbf{f}}(u,\lambda)) &= \mathbf{f}(u,\lambda) + \frac{N^2}{2T^2}\int_{-1/2}^{1/2} x^2 K_t(x)^2 dx \frac{\partial^2}{\partial u^2}\mathbf{f}(u,\lambda) \\ &\quad + \frac{1}{2M^2}\int_{-1/2}^{1/2} x^2 W(x)^2 dx \frac{\partial^2}{\partial \lambda^2}\mathbf{f}(u,\lambda) \\ &\quad + o\left(\frac{N^2}{T^2} + M^{-2}\right) + O\left(\frac{\log N}{N}\right), \end{aligned} \tag{14}$$

(iii)

$$\sum_{i,j=1}^{m} Var\left(\widehat{\mathbf{f}}_{i,j}(u,\lambda)\right) = \frac{M}{N}\sum_{i,j=1}^{m}\mathbf{f}_{i,j}(u,\lambda)^2\int_{-1/2}^{1/2}K_t(x)^2dx$$

$$\cdot\int_{-1/2}^{1/2}W(x)^2dx(2\pi+2\pi\{\lambda\equiv 0\ mod\pi\})+o\left(\frac{M}{N}\right). \tag{15}$$

Hence, we have

$$E\left\|\widehat{\mathbf{f}}(u,\lambda)-\mathbf{f}(u,\lambda)\right\|^2 = O\left(\frac{M}{N}\right)+O\left(M^{-2}+N^2T^{-2}\right)^2 = O\left(\frac{M}{N}\right), \tag{16}$$

where $\|A\|$ *is the Euclidean norm of the matrix* A*;* $\|A\| = \{tr\{AA^*\}\}^{1/2}$.

Lemma 2. *Assume that Assumptions 1-5 hold. Let* $\phi_j(u,\lambda)$*,* $j=1,\ldots,k$ *be* $m\times m$ *matrix-valued continuous function on* $[0,1]\times[-\pi,\pi]$ *which satisfies the same conditions as the transfer function matrix* $\mathbf{A}(u,\lambda)$ *in Assumption 2 and* $\phi_j(u,\lambda)^*=\phi_j(u,\lambda)$*,* $\phi_j(u,-\lambda)=\phi_j(u,\lambda)^*$*. Then*

$$L_T(\phi_j) = \sqrt{T}\Bigg\{\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}tr\left\{\phi_j\left(\frac{t}{T},\lambda\right)\mathbf{I}_N\left(\frac{t}{T},\lambda\right)\right\}d\lambda$$

$$-\int_0^1\int_{-\pi}^{\pi}tr\left\{\phi_j\left(u,\lambda\right)\mathbf{f}\left(u,\lambda\right)\right\}d\lambda du\Bigg\},\quad j=1,\ldots,k \tag{17}$$

have, asymptotically, a normal distribution with zero mean vector and covariance matrix V *whose* (i,j)*-the element is*

$$4\pi\int_0^1\Bigg[\int_{-\pi}^{\pi}tr\{\phi_i(u,\lambda)\mathbf{f}(u,\lambda)\phi_j(u,\lambda)\mathbf{f}(u,\lambda)\}d\lambda$$

$$+\frac{1}{4\pi^2}\sum_{a_1,a_2,a_3,a_4}\sum_{b_1,b_2,b_3,b_4}\kappa_{b_1,b_2,b_3,b_4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi_i(u,\lambda)_{a_1,a_2}\phi_j(u,\mu)_{a_4,a_3}$$

$$\cdot\mathbf{A}(u,\lambda)_{a_2,b_1}\mathbf{A}(u,-\lambda)_{a_1,b_2}\mathbf{A}(u,-\mu)_{a_4,b_3}\mathbf{A}(u,\mu)_{a_3,b_4}d\lambda d\mu\Bigg]du. \tag{18}$$

Assumption 5 does not coincide with Assumption A.1 (ii) of Dahlhaus (1997). As mentioned in A.3 Remarks of Dahlhaus (1997), Assumption A.1 (ii) of Dahlhaus (1997) is required because of the $\sqrt{T}$-unbiasedness at the boundary 0 and 1. If we assume that $\{\mathbf{X}_{2-N/2,T},\ldots,\mathbf{X}_{0,T}\}$ and $\{\mathbf{X}_{T+1,T},\ldots,\mathbf{X}_{T+N/2,T}\}$ are available with Assumption 5, then from Lemma 1 (i)

$$E\left(L_T(\phi_j)\right) = \sqrt{T}E\Bigg\{\frac{1}{T}\sum_{t=1}^{T}\int_{-\pi}^{\pi}tr\left\{\phi_j\left(\frac{t}{T},\lambda\right)\mathbf{I}_N\left(\frac{t}{T},\lambda\right)\right\}d\lambda$$

$$-\int_0^1\int_{-\pi}^{\pi}tr\left\{\phi_j\left(u,\lambda\right)\mathbf{f}\left(u,\lambda\right)\right\}d\lambda du\Bigg\}$$

$$= O\left(\sqrt{T}\left(\frac{N^2}{T^2}+\frac{\log N}{N}+\frac{1}{T}\right)\right) = o\left(1\right). \tag{19}$$

3. Measures of disparity. We suppose that we have a collection of zero-mean m-dimensional vector locally stationary time series $\mathbf{X}_{t,T} = (X_{t,T}^{(1)}, X_{t,T}^{(2)}, \ldots, X_{t,T}^{(m)})'$, $t = 1, 2, \ldots, T$. Denote by $p_i(\mathbf{x})$, $i = 1, 2$, the probability density functions of the $mT \times 1$ vector $\mathbf{x} = (\mathbf{X}_{1,T}', \mathbf{X}_{2,T}', \ldots, \mathbf{X}_{T,T}')'$ under two hypotheses Π_i, $i = 1, 2$, respectively. In the case of locally stationary processes, Π_i, $i = 1, 2$ are, respectively, described by the time varying spectral density matrices $\mathbf{f}_i(u, \lambda)$, $i = 1, 2$ corresponding to $mT \times mT$ matrices $\Sigma_T(p_i)$, $i = 1, 2$. Although theory developed later transcends the usual normal theory, it is convenient to use the normal assumption temporarily to motivate measures of disparity between the densities $p_i(\cdot)$, $i = 1, 2$.

One classical measure of disparity between two multivariate densities is the Kullback Leibler (KL) discrimination information, defined by

$$K(p_j; p_k) = E_p \left\{ \log \frac{p_j(\mathbf{x})}{p_k(\mathbf{x})} \right\}, \tag{20}$$

where E_p denotes the expectation under the density $p(\cdot)$. The KL discrimination information takes the form

$$K(p_j; p_k) = \frac{1}{2} \left(tr\{\Sigma_T(p_j)\Sigma_T(p_k)^{-1}\} - \log \frac{|\Sigma_T(p_j)|}{|\Sigma_T(p_k)|} - mT \right) \tag{21}$$

when $p_i(\mathbf{x})$ correspond to two hypothetical zero-mean multivariate normal distributions. The $mT \times mT$ covariance matrices $\Sigma_T(p_i)$ contain the $m \times m$ matrices $\Sigma_{s,t}^T(p_i)$, $s, t = 1, \ldots, T$ as blocks, where

$$\Sigma_{s,t}^T(p_i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(i\lambda(s-t)\right) \mathbf{A}_{s,T}^{\circ}(\lambda) \mathbf{\Omega} \mathbf{A}_{t,T}^{\circ}(-\lambda)' d\lambda. \tag{22}$$

Parzen (1990) proposed to use the Chernoff (CH) information measure

$$B_\alpha(p_j; p_k) = -\log E_{p_j} \left\{ \left(\frac{p_j(\mathbf{x})}{p_k(\mathbf{x})} \right)^{\alpha} \right\}, \tag{23}$$

as a measure of disparity between the two densities, where the measure is indexed by α, $0 < \alpha < 1$. For $\alpha = \frac{1}{2}$, the Chernoff information measure is the symmetric divergence measure. For two normal random vectors differing only in the covariance structure, the measure (23) takes the form

$$B_\alpha(p_j; p_k) = \frac{1}{2} \left(\log \frac{|\alpha \Sigma_T(p_j) + (1-\alpha)\Sigma_T(p_k)|}{|\Sigma_T(p_k)|} - \alpha \log \frac{|\Sigma_T(p_j)|}{|\Sigma_T(p_k)|} \right). \tag{24}$$

It is of interest to note the antisymmetry property $B_\alpha(p_j; p_k) = B_{1-\alpha}(p_k; p_j)$ and that $B_\alpha(p_j; p_k)$, scaled by $\alpha(1-\alpha)$ converges to $K(p_j; p_k)$ for $\alpha \to 0$ and to $K(p_k; p_j)$ for $\alpha \to 1$. Hence the Cernoff measure behaves like the two Kullback-Leibler measures for values of the parameter α that are near the boundaries 0 and 1.

It should be recognized that the information measures (21) and (24) both involve $mT \times mT$ matrices whose dimensions tend to infinity as $T \to \infty$. As in the case of stationary processes, it is natural to use spectral approximations in terms of the time varying spectral density matrices $\mathbf{f}_i(u, \lambda)$, $i = 1, 2$. The appropriate versions of (21) and (24) are

$$\begin{aligned} K(\mathbf{f}_j; \mathbf{f}_k) &= \lim_{T \to \infty} T^{-1} K(p_j; p_k) \\ &= \frac{1}{2} \int_0^1 \int_{-\pi}^{\pi} \left(tr\{\mathbf{f}_j(u, \lambda)\mathbf{f}_k^{-1}(u, \lambda)\} - \log \frac{|\mathbf{f}_j(u, \lambda)|}{|\mathbf{f}_k(u, \lambda)|} - m \right) \frac{d\lambda}{2\pi} du \end{aligned} \tag{25}$$

and

$$
\begin{aligned}
B_\alpha(\mathbf{f}_j;\mathbf{f}_k) &= \lim_{T\to\infty} T^{-1} B_\alpha(p_j;p_k) \\
&= \frac{1}{2}\int_0^1\int_{-\pi}^{\pi}\left(\log\frac{|\alpha\mathbf{f}_j(u,\lambda)+(1-\alpha)\mathbf{f}_k(u,\lambda)|}{|\mathbf{f}_k(u,\lambda)|} - \alpha\log\frac{|\mathbf{f}_j(u,\lambda)|}{|\mathbf{f}_k(u,\lambda)|}\right)\frac{d\lambda}{2\pi}du. \qquad (26)
\end{aligned}
$$

Note here that the time-varying spectral matrices $\mathbf{f}_i(u,\lambda)$ correspond to the multivariate densities $p_i(\mathbf{x})$. The advantage of (25) and (26) is that the evaluation problem is reduced to inverting $m\times m$ matrices. Both forms (25) and (26) are functions of the matrix product $\mathbf{f}_j(u,\lambda)\mathbf{f}_k^{-1}(u,\lambda)$ and can be generalized to the following disparity measure

$$
D_H(\mathbf{f}_j;\mathbf{f}_k) = \frac{1}{2}\int_0^1\int_{-\pi}^{\pi} H(\mathbf{f}_j(u,\lambda)\mathbf{f}_k^{-1}(u,\lambda))\frac{d\lambda}{2\pi}du \qquad (27)
$$

where $H(\cdot)$ is some matrix-valued function. To ensure that $D_H(\mathbf{f}_j;\mathbf{f}_k)$ has the quasi-distance property, we require $D_H(\mathbf{f}_j;\mathbf{f}_k)\ge 0$, and that the equality holds if and only if $\mathbf{f}_j=\mathbf{f}_k$ almost everywhere. The function $H(\mathbf{Z})$ must have a unique minimum at $\mathbf{Z}=\mathbf{E}_m$, the identity matrix. There are many possible choices of $H(\mathbf{Z})$ such that $D_H(\cdot;\cdot)$ satisfies the quasi-distance property, but we consider only the two cases corresponding to (25) and (26):

$$
H_K(\mathbf{Z}) = tr\{\mathbf{Z}\} - \log|\mathbf{Z}| - m \qquad (28)
$$

and

$$
H_{B_\alpha}(\mathbf{Z}) = \log|\alpha\mathbf{Z}+(1-\alpha)\mathbf{E}_m| - \alpha\log|\mathbf{Z}|. \qquad (29)
$$

Note that another possible choice is the quadratic function

$$
H_Q(\mathbf{Z}) = \frac{1}{2}tr\left(\mathbf{Z}-\mathbf{E}_m\right)^2. \qquad (30)
$$

Generally, $D_H(\cdot;\cdot)$ is not symmetric but can easily be made so by defining

$$
\widetilde{H}(\mathbf{Z}) = H(\mathbf{Z}) + H(\mathbf{Z}^{-1}). \qquad (31)
$$

The general form (27) can be approximated by sums over frequencies of the form $\lambda_s = 2\pi s/T$ and time $u_t = t/T$, $s,t=1,2,\dots T$, i.e.,

$$
D_H(\mathbf{f}_j;\mathbf{f}_k) \approx \frac{1}{2}T^{-2}\sum_{s,t=1}^{T} H\left(\mathbf{f}_j(u_t,\lambda_s)\mathbf{f}_k^{-1}(u_t,\lambda_s)\right). \qquad (32)
$$

4. Discriminant analysis. Suppose that we wish to investigate the problem of classifying a realization $\mathbf{X}_T = (\mathbf{X}'_{2-N/2,T},\dots,\mathbf{X}'_{1,T},\dots,\mathbf{X}'_{T,T},\dots,\mathbf{X}'_{T+N/2,T})'$ into one of two known categories Π_j, $j=1,2$, where Π_j is described by the time varying spectral density matrix $\mathbf{f}_j(u,\lambda)$. Let $\widehat{\mathbf{f}}_T(u,\lambda)$ be the nonparametric time varying spectral density estimator given by (9), which is based on observation to be classified. We measure the disparity between the sample spectrum of $\mathbf{X}_T$ and category Π_j by $D_H(\widehat{\mathbf{f}}_T;\mathbf{f}_j)$. Then the proposed rule is to classify $\mathbf{X}_T$ into Π_1 or Π_2 according to $D_H>0$ or $D_H\le 0$, where D_H is the discriminant function defined by

$$
D_H = D_H(\widehat{\mathbf{f}}_T;\mathbf{f}_2) - D_H(\widehat{\mathbf{f}}_T;\mathbf{f}_1). \qquad (33)
$$

In this section we examine the asymptotic properties of discriminant function (33). Assume that the category Π_j is an m-variate linear process of the form $\mathbf{X}_{t,T} = \sum_{k=-\infty}^{\infty}\mathbf{a}_{t,T}^{(j)}(k)\varepsilon_{t-k}$,

where $m \times m$ matrices $\mathbf{a}_{t,T}^{(j)}(k)$'s and i.i.d. $m \times 1$ zero mean vectors ε_t's satisfy Assumptions 1 and 2. The use of $\widehat{\mathbf{f}}_T(u,\lambda)$ instead of the data tapered periodogram $\mathbf{I}_N(u,\lambda)$ is essential, because $D_H(\mathbf{I}_N;\mathbf{g})$ does not converge in probability to $D_H(\mathbf{f}_j;\mathbf{g})$ under Π_j if $D_H(\mathbf{I}_N;\mathbf{g})$ is nonlinear with respect to $\mathbf{I}_N$ (See Taniguchi and Kakizawa (2000)).
We discuss the performance of the discriminant function (33). First, we evaluate the asymptotics of the misclassification probabilities based on D_H;

$$P_{D_H}(2|1) = \Pr(D_H \le 0|\Pi_1) \tag{34}$$

and

$$P_{D_H}(1|2) = \Pr(D_H > 0|\Pi_2). \tag{35}$$

It is shown that $P_{D_H}(2|1)$ and $P_{D_H}(1|2)$ converge to zero as $T \to \infty$ if $\mathbf{f}_1(u,\lambda) \neq \mathbf{f}_2(u,\lambda)$. Next assuming that Π_1 is contiguous to π_2, the limit of the two misclassification probabilities is evaluated. Then we will elucidate the asymptotic optimality and robustness
In what follows, set $(j,k) = (1,2)$ or $(2,1)$. We give the following assumption on matrix-valued function $H(\mathbf{Z})$.

Assumption 6. **(i)** *$H : \mathbb{C}^{m\times m} \to \mathbb{R}$ is real-valued holomorphic function defined on an open set D in $\mathbb{C}^{m\times m}$.*

(ii) *$H(\mathbf{E}_m) = 0$ and $H^{(1)}(\mathbf{E}_m) = \mathbf{0}$ ($m \times m$ zero matrix), where $H^{(1)}(\{\cdot\})$ is the first derivative of $\mathbf{Z}$ at $\{\cdot\}$ whose (a,b)-th element is $\frac{\partial}{\partial Z_{ab}}H(\mathbf{Z})$. The $m^2 \times m^2$ Hessian matrix of $H(\mathbf{Z})$, defined by*

$$\frac{\partial}{\partial(vec\ \mathbf{Z})'}\left(\frac{\partial}{\partial(vec\ \mathbf{Z})'}H(\mathbf{Z})\right)' \tag{36}$$

is positive definite at $\mathbf{Z} = \mathbf{E}_m$. That is, $H(\mathbf{Z})$ has a unique minimum zero at $\mathbf{Z} = \mathbf{E}_m$.

(iii) *The $m \times m$ matrix $\mathbf{Q}_{j,k}(u,\lambda)$ defined by*

$$\mathbf{Q}_{j,k}(u,\lambda) = \mathbf{f}_k^{-1}(u,\lambda)\left[H^{(1)}\left(\mathbf{f}_j(u,\lambda)\mathbf{f}_k^{-1}(u,\lambda)\right)\right]' \tag{37}$$

satisfies $\mathbf{Q}_{j,k}(u,\lambda)^* = \mathbf{Q}_{j,k}(u,\lambda)$ and $\mathbf{Q}_{j,k}(u,-\lambda) = \mathbf{Q}_{j,k}(u,\lambda)'$.

Theorem 1. *Under the Assumptions 1-6, suppose that $\mathbf{f}_1(u,\lambda) \neq \mathbf{f}_2(u,\lambda)$ on a set of positive Lebesgue measure. Then under Π_j, $D_H \xrightarrow{P} (-1)^{j+1}D_H(\mathbf{f}_j;\mathbf{f}_k)$ and*

$$\sqrt{T}\{D_H + (-1)^j D_H(\mathbf{f}_j;\mathbf{f}_k)\} \xrightarrow{D} N(0, V_H^2(j,k)), \quad as\ T \to \infty, \tag{38}$$

where $D_H(\mathbf{f}_j;\mathbf{f}_k)$ is the integral disparity (27) and

$$V_H^2(j,k) = \int_0^1 \left[\frac{1}{4\pi}\int_{-\pi}^{\pi} tr\{\mathbf{Q}_{j,k}(u,\lambda)\mathbf{f}_j(u,\lambda)\}^2 d\lambda \right.$$
$$\left. + \frac{1}{64\pi^4}\sum_{a,b,c,d}\kappa_{a,b,c,d}\gamma_{a,b}^{(j,k)}(u)\gamma_{c,d}^{(j,k)}(u)\right] du, \tag{39}$$

with

$$\mathbf{\Gamma}_H^{(j,k)}(u) = \left\{\gamma_{a,b}^{(j,k)}(u)\right\}_{a,b=1,\dots,m} = \int_{-\pi}^{\pi}\mathbf{A}_j^*(u,\lambda)\mathbf{Q}_{j,k}(u,\lambda)\mathbf{A}_k(u,\lambda)d\lambda. \tag{40}$$

In view of Theorem 1, the limiting forms of misclassification probabilities (34) and (35) satisfy $\lim_{T\to\infty} P_{D_H}(k|j) = 0$ for $(j,k) = (1,2),(2,1)$. This shows that the discriminant D_H is consistent. From (39), one may also approximate them as the normal integrals

$$P_{D_H}(k|j) \approx \Phi\left(-\sqrt{T}\frac{D_H(\mathbf{f}_j;\mathbf{f}_k)}{V_H(j,k)}\right), \tag{41}$$

which depend on the fourth-order cumulants unless (40) is a zero matrix. To look at robustness, we assume that the hypothetical m-variate linear process is generated by

$$\mathbf{X}_{t,T} = \sum_{s=-\infty}^{\infty} \mathbf{a}_{t,T}^{(1)}(s)\varepsilon_{t-s} \tag{42}$$

under Π_1 and by

$$\mathbf{X}_{t,T} = \sum_{s=-\infty}^{\infty} \left\{\mathbf{a}_{t,T}^{(1)}(s) + T^{-1/2}\mathbf{a}_{t,T}^{(2)}(s)\right\}\varepsilon_{t-s} \tag{43}$$

under Π_2. Thus, the time varying spectral densities associated with Π_1 and Π_2 are

$$\mathbf{f}_1(u,\lambda) = \mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^* \tag{44}$$

and

$$\begin{aligned}\mathbf{f}_2(u,\lambda) = &\left\{\mathbf{A}^{(1)}(u,\lambda) + T^{-1/2}\mathbf{A}^{(2)}(u,\lambda)\right\}\mathbf{\Omega}\\ &\left\{\mathbf{A}^{(1)}(u,\lambda) + T^{-1/2}\mathbf{A}^{(2)}(u,\lambda)\right\}^*,\end{aligned} \tag{45}$$

with $\mathbf{A}^{(i)}(u,\lambda) = \sum_{s=-\infty}^{\infty} \mathbf{a}_s^{(i)}(u)\exp\{-i\lambda s\}$, $i = 1,2$. The quantities $D_H(\mathbf{f}_j;\mathbf{f}_k)$ and $V_H(j,k)$ are determined by the local property of the function $H(\mathbf{Z})$ at $\mathbf{Z} = \mathbf{E}_m$.

Assumption 7. *The $m^2 \times m^2$ Hessian matrix of $H(\mathbf{Z})$ at $\mathbf{E}_m$ is $c\mathbf{K}_m$, where $\mathbf{K}_m$ is the commutation matrix (e.g., Magnus and Neudecker (1988)) and $c > 0$.*

Note that H_K, H_{B_α} and H_Q in (28), (29) and (30) satisfy Assumptions 6 and 7.

Theorem 2. *Let $\mathbf{f}_1$ and $\mathbf{f}_2$, defined by (44) and (45), be the hypothetical time varying hypothetical spectral density matrices of m-variate linear processes (42) and (43), respectively. Under Assumptions 1-7, if $\mathbf{a}_0^{(2)}(u) = \mathbf{0}$ ($m \times m$ zero matrix), the asymptotic misclassification probabilities are independent of the non-Gaussinanity of the process, and are given by*

$$\begin{aligned}\lim_{T\to\infty} P_{D_H}(2|1) &= \lim_{T\to\infty} P_{D_H}(1|2)\\ &= \Phi\left(-\frac{1}{2}\sqrt{\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\{\mathbf{\Delta}(u,\lambda)\}^2 d\lambda du}\right),\end{aligned} \tag{46}$$

with

$$\begin{aligned}\mathbf{\Delta}(u,\lambda) = &\{\mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(2)}(u,\lambda)^* + \mathbf{A}^{(2)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^*\}\\ &\{\mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^*\}^{-1}.\end{aligned} \tag{47}$$

Furthermore, if the process concerned is Gaussian, the exact Gaussian log-likelihood ratio is

$$\Lambda_T(p_1,p_2) \equiv \frac{1}{T}\ \textit{Gaussian log likelihood ratio}$$
$$= \frac{1}{2T}\mathbf{X}_T'\left(\Sigma^T(\mathbf{p}_1)^{-1} - \Sigma^T(\mathbf{p}_2)^{-1}\right)\mathbf{X}_T - \frac{1}{2T}\log\frac{|\Sigma^T(\mathbf{p}_2)|}{|\Sigma^T(\mathbf{p}_1)|}. \tag{48}$$

According to Proposition 2.5 and Lemma A.8 of Dahlhaus (2000), it is seen that, under Π_j, for each $\varepsilon > 0$,

$$\begin{aligned}
&E\left(\sqrt{T}\Lambda_T(p_1,p_2)\right)\\
&= \frac{\sqrt{T}}{2T}\left\{tr\left\{\Sigma^T(\mathbf{p}_j)\left(\Sigma^T(\mathbf{p}_1)^{-1} - \Sigma^T(\mathbf{p}_2)^{-1}\right)\right\} - \log\frac{|\Sigma^T(\mathbf{p}_2)|}{|\Sigma^T(\mathbf{p}_1)|}\right\}\\
&= \frac{\sqrt{T}}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left\{tr\left\{\mathbf{f}_j(u,\lambda)\left(\mathbf{f}_2^{-1}(u,\lambda) - \mathbf{f}_1^{-1}(u,\lambda)\right)\right\} - \log\frac{|\mathbf{f}_2(u,\lambda)|}{|\mathbf{f}_1(u,\lambda)|}\right\}d\lambda du\\
&\quad + O(T^{-\frac{1}{2}+\varepsilon} + T^{-\frac{1}{2}}\log^{11}T)\\
&= (-1)^{j+1}\sqrt{T}D_{H_K}(\mathbf{f}_j;\mathbf{f}_k) + o(1)
\end{aligned} \tag{49}$$

and

$$\begin{aligned}
Var\left(\sqrt{T}\Lambda_T(p_1,p_2)\right) &= \frac{1}{2T}tr\left\{\Sigma^T(\mathbf{p}_j)\left(\Sigma^T(\mathbf{p}_1)^{-1} - \Sigma^T(\mathbf{p}_2)^{-1}\right)\right\}^2\\
&= \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{f}_j(u,\lambda)\left(\mathbf{f}_2^{-1}(u,\lambda) - \mathbf{f}_1^{-1}(u,\lambda)\right)\right\}^2 d\lambda du\\
&\quad + O(T^{-1}\log^{23}T).
\end{aligned} \tag{50}$$

Therefore,

$$\begin{aligned}
&\lim_{T\to\infty} P_{GLR}(2|1) = \lim_{T\to\infty} P_{GLR}(1|2)\\
&= \lim_{T\to\infty} P_{D_H}(2|1) = \lim_{T\to\infty} P_{D_H}(1|2)\\
&= \Phi\left(-\frac{1}{2}\sqrt{\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\{\boldsymbol{\Delta}(u,\lambda)\}^2 d\lambda du}\right),
\end{aligned} \tag{51}$$

where

$$P_{GLR}(2|1) = Pr(\Lambda_T(p_1,p_2) \le 0|\Pi_1) \tag{52}$$

and

$$P_{GLR}(1|2) = Pr(\Lambda_T(p_1,p_2) > 0|\Pi_2), \tag{53}$$

that is, the discriminant criterion based on D_H is asymptotically Gaussian optimal.

Remark 1. ***Peak Robustness of*** B_α***.***
We consider a case where the time varying spectral density of $\mathbf{X}_T$ *is contaminated by a sharp peak. In this case, we can see that* $B_\alpha(\mathbf{f}_j;\mathbf{f}_k)$ *is robust with respect to peak, but* $K(\mathbf{f}_j;\mathbf{f}_k)$ *is not so. Define*

$$\bar{\mathbf{f}}_i(u,\lambda) = \begin{cases} \mathbf{f}_i(u,\lambda) & \textit{on } \Omega = [-\pi,\pi] - \Omega_\epsilon;\\ \mathbf{f}_i(u,\lambda)/\epsilon^r & \textit{on } \Omega_\epsilon, \end{cases} \tag{54}$$

where $\Omega_\epsilon = [\lambda_0, \lambda_0 + \epsilon]$ *is an interval in* $[-\pi, \pi]$ *for sufficiently small* $\epsilon > 0$ *and* $r > 1$. *Suppose that* $\mathbf{f}_1(u,\lambda) \not\equiv \mathbf{f}_2(u,\lambda)$ *on a set of a positive Lebesgue measure. Then, under Assumption 2, it can be shown that*

$$\lim_{\epsilon \to 0} |B_\alpha(\bar{\mathbf{f}}_j, \mathbf{f}_j, \mathbf{f}_k) + (-1)^j B_\alpha(\mathbf{f}_j, \mathbf{f}_k)| = 0 \quad for \quad \alpha \in (0,1), \tag{55}$$

$$\lim_{\epsilon \to 0} |K(\bar{\mathbf{f}}_j, \mathbf{f}_j, \mathbf{f}_k) + (-1)^j K(\mathbf{f}_j, \mathbf{f}_k)| = \infty \tag{56}$$

where

$$B_\alpha(\bar{\mathbf{f}}_j, \mathbf{f}_j, \mathbf{f}_k) = B_\alpha(\bar{\mathbf{f}}_j, \mathbf{f}_k) - B_\alpha(\bar{\mathbf{f}}_j, \mathbf{f}_j) \tag{57}$$

and

$$\begin{aligned} &K(\bar{\mathbf{f}}_j, \mathbf{f}_j, \mathbf{f}_k) = K(\bar{\mathbf{f}}_j, \mathbf{f}_k) - K(\bar{\mathbf{f}}_j, \mathbf{f}_j) \\ &= \frac{1}{4\pi} \int_0^1 \int_{-\pi}^{\pi} \left\{ \log \frac{|\mathbf{f}_k(u,\lambda)|}{|\mathbf{f}_j(u,\lambda)|} + tr \left\{ \bar{\mathbf{f}}_j(u,\lambda) \left(\mathbf{f}_k^{-1}(u,\lambda) - \mathbf{f}_j^{-1}(u,\lambda) \right) \right\} \right\} d\lambda du. \end{aligned} \tag{58}$$

That is, $B_\alpha(\mathbf{f}_j; \mathbf{f}_k)$ *is insensitive to sharp peak in the spectral density, while* $K(\mathbf{f}_j; \mathbf{f}_k)$ *is sensitive. Thus, the discriminant statistic* $B_\alpha(\mathbf{f}_j; \mathbf{f}_k)$ *is better than* $K(\mathbf{f}_j; \mathbf{f}_k)$ *if the time varying spectral density of* $\mathbf{X}_T$ *is contaminated by a sharp peak.*

Numerical example. *We consider the time varying AR(2) model*

$$X_{t,T} = b_\theta^{(1)} \left(\frac{t}{T} \right) X_{t-1,T} + b_\theta^{(2)} \left(\frac{t}{T} \right) X_{t-2,T} + \varepsilon_t, \tag{59}$$

with $b_\theta^{(i)}(u) = \theta^{(i)} \exp(-2\theta^{(i)}(u+1))$, $i = 1, 2$, *and* ε_t *are i.i.d. random variables with probability density*

$$p(x) = \exp\left(-(x+1)\right), \quad x > -1. \tag{60}$$

Then the time varying spectral density is given by

$$f_\theta(u,\lambda) = \frac{1}{2\pi} \left| \frac{1}{1 - \theta^{(1)} \exp(-2\theta^{(1)}(u+1)) \exp(i\lambda) - \theta^{(2)} \exp(-2\theta^{(2)}(u+1)) \exp(2i\lambda)} \right|^2. \tag{61}$$

Now, we give two examples for the model (59);

Example 1 $(\theta^{(1)}, \theta^{(2)}) = (0.8, 0):$ *AR(1) model.*

Example 2 $(\theta^{(1)}, \theta^{(2)}) = (0.8, 0.2):$ *AR(2) model.*

In Figure 1, AR(1) coefficient $b_\theta^{(1)}(u)$ *(real line) and AR(2) coefficient* $b_\theta^{(2)}(u)$ *(dotted line) of Example 2 are plotted. From this figure, it is seen that* $b_\theta^{(1)}(u)$ *decreases as* $u \to 1$, *on the other hand* $b_\theta^{(2)}(u)$ *is almost constant.*

Figures 1 is about here.

For Examples 1 and 2, the observed stretch $\{X_{t,T}\}$ *with* $T = 2^8$ *are plotted in Figures 2 and 3, respectively. The time varying spectral densities of them are, respectively, given in Figures 4 and 5.*

Figures 2 and 3 are about here.

Figures 4 and 5 are about here.

Next, in Example 1, let the time varying spectral densities associated with Π_1 *and* Π_2 *be*

$$\begin{cases} \Pi_1: \quad f_1(u,\lambda) = \frac{1}{2\pi}\left|\frac{1}{1-\theta_1^{(1)}\exp(-2\theta_1^{(1)}(u+1))\exp(i\lambda)}\right|^2 \\ \Pi_2: \quad f_2(u,\lambda) = \frac{1}{2\pi}\left|\frac{1}{1-\theta_2^{(1)}\exp(-2\theta_2^{(1)}(u+1))\exp(i\lambda)}\right|^2, \end{cases} \tag{62}$$

respectively. We consider the following two cases;

CASE *I The time varying spectral density of process concerned is not contaminated by a sharp peak, That is, we actually observe*

$$Y_{t,T} = X_{t,T}. \tag{63}$$

CASE *II The time varying spectral density of process concerned is contaminated by a sharp peak. That is, we actually observe*

$$Y_{t,T} = X_{t,T}^1 + X_{t,T}^2, \tag{64}$$

where

$$X_{t,T}^1 = X_{t,T} \tag{65}$$

and

$$\begin{aligned} X_{t,T}^2 &= \int_{-\pi}^{\pi}\left\{\epsilon^{-r/2}-1\right\} A_\theta\left(\frac{t}{T},\lambda\right)\exp(i\lambda t) I_{(\lambda_0,\lambda_0+\epsilon]} d\xi(\lambda) \\ &\approx \left\{\epsilon^{-r/2}-1\right\} A_\theta\left(\frac{t}{T},\lambda_0\right)\exp(i\lambda_0 t)\left(\xi(\lambda_o+\epsilon)-\xi(\lambda_0)\right) \\ &\approx \left\{\epsilon^{-r/2}-1\right\} A_\theta\left(\frac{t}{T},\lambda_0\right)\exp(i\lambda_0 t) \\ &\quad \cdot\frac{1}{2\pi}\sum_{j=-n}^{n}\varepsilon_j\frac{\exp(-ij(\lambda_0+\epsilon))-\exp(-ij\lambda_0)}{-ij}. \end{aligned} \tag{66}$$

For the cases II, the observed stretch $\{Y_{t,T}\}$ *with* $T = 2^8$, $\theta^{(1)} = 0.8$, $\lambda_0 = -\pi/4$, $r = 3$ *and* $\epsilon = 0.01$ *are plotted in Figures 6. The time varying spectral density* $\bar{f}_1(u,\lambda)$ *are given in Figures 7.*

Figures 6 and 7 are about here.

Finally, in Figures 8 and 9, we plot the pair graphs $K(\bar{f}_1, f_1, f_2)$, $K(f_1, f_2)$ *and* $B_{1/2}(\bar{f}_1, f_1, f_2)$, $B_{1/2}(f_1, f_2)$ *with* $(\theta_1^{(1)}, \theta_2^{(1)}) = (0.8, 0.9)$, *respectively. From Figures 8 and 9, it is seen that* $K(\bar{f}_1, f_1, f_2)$ *diverges as* $\epsilon \to 0$, *on the other hand* $B_{1/2}(\bar{f}_1, f_1, f_2)$ *converges to* $B_{1/2}(f_1, f_2)$ *as* ϵ *tends to* 0.

Figures 8 and 9 are about here.

5. Asymptotically non-Gaussian optimal classification. Suppose that $\mathbf{X}_T = (X_{1,T}, \ldots, X_{T,T})'$ is a realization of a scalar-valued locally stationary process with transfer function A_θ° where the corresponding A_θ is uniformly bounded from above and below, and time varying spectral density $f_\theta(u,\lambda) := |A_\theta(u,\lambda)|^2$ depends on a parameter vector $\theta = (\theta_1, \ldots, \theta_q) \in \Theta \subset \mathbb{R}^q$.
Let Π_1 and Π_2 be two categories with probability density functions $p_1(\mathbf{X})$ and $p_2(\mathbf{X})$, respectively. We investigate the problems of classifying a locally stationary process $\{\mathbf{X}_T\}$ into one of two categories described by two hypotheses:

$$\begin{aligned} &\Pi_1 : f_1(u,\lambda) = f\,(u,\lambda|\theta) \\ &\Pi_2 : f_2(u,\lambda) = f\left(u,\lambda|\theta_T = \theta + \frac{h}{\sqrt{T}}\right), \end{aligned} \tag{67}$$

where $\theta \in \Theta \subset \mathbb{R}^q$ and $h = (h_1, \ldots, h_q)'$. We assign the observed stretch $\mathbf{X}_T$ to category Π_1 if $\mathbf{X}_T$ falls into region R_1; otherwise we assign it to Π_2, where R_1 and R_2 are exclusive and exhaustive regions in $\mathbb{R}^T$. It is well known that the classification regions defined by

$$R_1 = \left[\mathbf{X}_T : \Lambda_T(p_1,p_2) = \log\frac{p_1(\mathbf{X}_T)}{p_2(\mathbf{X}_T)} > 0\right] \tag{68}$$

give the optimal classification (See Anderson (1984)).
Introducing the notations $\nabla_i = \frac{\partial}{\partial\theta_i}$, $\nabla = (\nabla_1, \ldots, \nabla_q)'$, $\nabla_{ij} = \frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}$, $\nabla^2 = (\nabla_{ij})_{ij=1,\ldots,q}$, we make the following assumption.

Assumption 8. (A1) *There exists a constant K with*

$$\sup_{t,\lambda}\left|\nabla^s\{A^\circ_{\theta,t,T} - A_\theta(\frac{t}{T},\lambda)\}\right| \le KT^{-1} \tag{69}$$

for $s = 0,1,2$. The components of $A_\theta(u,\lambda)$, $\nabla A_\theta(u,\lambda)$ and $\nabla^2 A_\theta(u,\lambda)$ are differentiable in u and λ with uniformly continuous derivatives $\frac{\partial}{\partial u}\frac{\partial}{\partial\lambda}$.

Writing

$$\varepsilon_t = \int_{-\pi}^{\pi} \exp(i\lambda t)d\xi(\lambda), \tag{70}$$

we assume the followings.

Assumption 9. (B1) *ε_t's are i.i.d. random variables with mean 0, variance 1 and finite fourth order moment $E(\varepsilon_t^4)$. Furthermore the distribution is absolutely continuous with respect to Lebesgue measure, and has the probability density $p(z) > 0$ on $\mathbb{R}$.*

(B2) *$p(\cdot)$ satisfies*

$$\lim_{|z|\to\infty} p(z) = 0, \quad \text{and} \quad \lim_{|z|\to\infty} zp(z) = 0. \tag{71}$$

(B3) *The continuous derivatives Dp, $D^2p \equiv D(Dp)$ and $D^3p \equiv D(D^2p)$ of $p(\cdot)$ exist on $\mathbb{R}$, and D^3p satisfies the Lipschitz condition.*

(B4)

$$\mathcal{F}(p) = \int (\phi(z))^2 p(z)dz < \infty, \tag{72}$$

$$\int (D^s\phi(z))^2 p(z)dz < \infty, \quad s = 1, 2, \tag{73}$$

$$E(\varepsilon_t \phi^2(\varepsilon_t)), \quad E(\varepsilon_t^2 \phi^2(\varepsilon_t)) \quad \textit{and} \quad E(\phi(z))^4 < \infty \tag{74}$$

and

$$\int D^2 p(z)dz = 0, \quad \lim_{|z|\to\infty} Dp(z)z^2 = 0, \tag{75}$$

where $\phi(z) = \frac{Dp(z)}{p(z)}$.

Assumption 10. **(C1)** $\{X_{t,T}\}$ *has the* $MA(\infty)$ *and* $AR(\infty)$ *representations*

$$X_{t,T} = \sum_{j=0}^{\infty} a^{\circ}_{\theta,t,T}(j)\varepsilon_{t-j}, \tag{76}$$

$$a^{\circ}_{\theta,t,T}(0)\varepsilon_t = \sum_{k=0}^{\infty} b^{\circ}_{\theta,t,T}(k) X_{t-k,T}, \tag{77}$$

where $a^{\circ}_{\theta,t,T}(j)$, $b^{\circ}_{\theta,t,T}(k) \in \mathbb{R}$, $b^{\circ}_{\theta,t,T}(0) \equiv 1$ *and* $a^{\circ}_{\theta,t,T}(j) = a^{\circ}_{\theta,0,T}(j) = a^{\circ}_{\theta}(j)$ *for* $t \leq 0$.

(C2) *Every* $a^{\circ}_{\theta,t,T}(j)$ *is continuously three times differentiable with respect to* θ*, and the derivatives satisfy*

$$\sup_{t,T}\left\{\sum_{j=0}^{\infty}(1+j)|\nabla_{i_1}\cdots\nabla_{i_s} a^{\circ}_{\theta,t,T}(j)|\right\} < \infty \quad \textit{for } s = 0, 1, 2, 3. \tag{78}$$

(C3) *Every* $b^{\circ}_{\theta,t,T}(k)$ *is continuously three times differentiable with respect to* θ*, and the derivatives satisfy*

$$\sup_{t,T}\left\{\sum_{k=0}^{\infty}(1+k)|\nabla_{i_1}\cdots\nabla_{i_s} b^{\circ}_{\theta,t,T}(k)|\right\} < \infty \quad \textit{for } s = 0, 1, 2, 3. \tag{79}$$

(C4)

$$a^{\circ}_{\theta,t,T}(0) = \exp\left\{\frac{1}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(f^{\circ}_{\theta,t,T}(\lambda))d\lambda\right\}, \tag{80}$$

where $f^{\circ}_{\theta,t,T}(\lambda) = |A^{\circ}_{\theta,t,T}(\lambda)|^2$.

By (76) and (77) we have

$$a^{\circ}_{\theta,t,T}(0)\varepsilon_t = \sum_{k=0}^{t-1} b^{\circ}_{\theta,t,T}(k) X_{t-k,T} + \sum_{r=0}^{\infty} c^{\circ}_{\theta,t,T}(r)\varepsilon_{-r}, \tag{81}$$

where

$$c^{\circ}_{\theta,t,T}(r) = \sum_{s=0}^{r} b^{\circ}_{\theta,t,T}(t+s) a^{\circ}_{\theta}(r-s). \tag{82}$$

From Assumption 10 it follows that

$$\sum_{r=0}^{\infty}|c^{\circ}_{\theta,t,T}(r)| \le \sum_{k=t}^{\infty}\sum_{l=0}^{\infty}|b^{\circ}_{\theta,t,T}(k)||a^{\circ}_{\theta}(l)|$$

$$\le \sum_{l=0}^{\infty}|a^{\circ}_{\theta}(l)|\frac{1}{t}\sum_{k=t}^{\infty}k|b^{\circ}_{\theta,t,T}(k)| = O(t^{-1}). \tag{83}$$

By Theorem 1 of Hirukawa and Taniguchi (2004), we have for all $\theta \in \Theta$, under Π_1, as $T \to \infty$, the log-likelihood ratio $\Lambda_T(p_1,p_2)$ has, asymptotically, normal distribution $\mathcal{N}(\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h)$, where

$$\Gamma(\theta) = \int_0^1 \left\{ \frac{\mathcal{F}(p)}{4\pi}\int_{-\pi}^{\pi}\frac{(\nabla f_\theta(u,\lambda))(\nabla f_\theta(u,\lambda))'}{|f_\theta(u,\lambda)|^2}d\lambda \right.$$

$$+\frac{1}{16\pi^2}\left\{E(\varepsilon_t^2\phi^2(\varepsilon_t)) - 2\mathcal{F}(p) - 1\right\}$$

$$\left. \cdot\left\{\int_{-\pi}^{\pi}\frac{\nabla f_\theta(u,\lambda)}{f_\theta(u,\lambda)}d\lambda\right\}\left\{\int_{-\pi}^{\pi}\frac{\nabla f_\theta(u,\lambda)}{f_\theta(u,\lambda)}d\lambda\right\}'\right\}du. \tag{84}$$

Furthermore, under Π_2, $\Lambda_T(p_1,p_2) \xrightarrow{\mathcal{D}} \mathcal{N}(-\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h)$, hence

$$\lim_{T\to\infty}P_{LR}(2|1) = \lim_{T\to\infty}P_{LR}(1|2)$$

$$= \Phi\left(-\frac{1}{2}\sqrt{h'\Gamma(\theta)h}\right), \tag{85}$$

where

$$P_{LR}(2|1) = Pr(\Lambda_T(p_1,p_2) \le 0|\Pi_1) \tag{86}$$

and

$$P_{LR}(1|2) = Pr(\Lambda_T(p_1,p_2) > 0|\Pi_2), \tag{87}$$

Since $(\varepsilon_s, s \le 0)$ are unobservable, instead of $\Lambda_T(p_1,p_2)$ we use the "quasi-log-likelihood ratio"

$$F_T(p_1,p_2) = \log\frac{F_T(\theta)}{F_T(\theta_T)}, \tag{88}$$

with

$$F_T(\theta) = \prod_{t=1}^{T}\frac{1}{a^{\circ}_{\theta,t,T}(0)}p\left\{\frac{\sum_{k=0}^{t-1}b^{\circ}_{\theta,t,T}(k)X_{t-k,T}}{a^{\circ}_{\theta,t,T}(0)}\right\}, \tag{89}$$

for classification criterion.

Theorem 3. *The discriminant criterion based on the quasi-log-likelihood ratio is asymptotically optimal.*

Since hypotheses are often unknown, when there are training samples, the results may also be extended to the plug-in version of D_H, obtained by replacing $\mathbf{f}_j$ with the estimator $\widetilde{\mathbf{f}}_j$ based on the training samples.

6. Proofs. This section provides the proofs of theorems.

Proof of Theorem 1. Let

$$\widehat{H}_j(u,\lambda) \equiv H\left(\widehat{\mathbf{f}}_T(u,\lambda)\mathbf{f}_k^{-1}(u,\lambda)\right) - H\left(\mathbf{f}_j(u,\lambda)\mathbf{f}_k^{-1}(u,\lambda)\right)$$
$$-tr\left\{\mathbf{Q}_{j,k}(u,\lambda)(\widehat{\mathbf{f}}_T(u,\lambda)-\mathbf{f}_j(u,\lambda))\right\}, \tag{90}$$

then from Lemma 1, the same argument as in Theorem 1 of Taniguchi et al. (1996), leads to, under Π_j

$$\widehat{H}_j(u,\lambda) = O_P\left(\frac{M}{N}\right) \tag{91}$$

and

$$H\left(\widehat{\mathbf{f}}_T(u,\lambda)\mathbf{f}_j^{-1}(u,\lambda)\right) = O_P\left(\frac{M}{N}\right). \tag{92}$$

uniformly in λ and u. Since, D_H is written as

$$D_H = \frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi}\left[H\left(\widehat{\mathbf{f}}_T(u,\lambda)\mathbf{f}_2^{-1}(u,\lambda)\right) - H\left(\widehat{\mathbf{f}}_T(u,\lambda)\mathbf{f}_1^{-1}(u,\lambda)\right)\right]d\lambda du, \tag{93}$$

it follows from (91) and (92), under Π_j

$$\begin{aligned}
&\sqrt{T}\{D_H + (-1)^j D_H(\mathbf{f}_j;\mathbf{f}_k)\}\\
&= \frac{(-1)^{j+1}\sqrt{T}}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\lambda)(\widehat{\mathbf{f}}_T(u,\lambda)-\mathbf{f}_j(u,\lambda))\right\}d\lambda du + o_P(1)\\
&= \frac{(-1)^{j+1}}{4\pi}S_T + o_P(1) \quad (say).
\end{aligned}\tag{94}$$

According to Lemma 2,

$$\begin{aligned}
&L_T\left(\frac{(-1)^{j+1}}{4\pi}\mathbf{Q}_{j,k}\right)\\
&= \frac{(-1)^{j+1}\sqrt{T}}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\lambda)(\mathbf{I}_N(u,\lambda)-\mathbf{f}_j(u,\lambda))\right\}d\lambda du + O(T^{-\frac{1}{2}})\\
&= \frac{(-1)^{j+1}}{4\pi}L_T + o(1) \quad (say)
\end{aligned}\tag{95}$$

have, asymptotically, a normal distribution with zero mean vector and covariance matrix $V_H^2(j,k)$. Thus, the proof of Theorem 1 is complete if we show $S_T - L_T = o_p(1)$. From the definition of $\widehat{\mathbf{f}}$, it follows that

$$\begin{aligned}
&S_T - L_T\\
&= \sqrt{T}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\lambda)\left(\int_{-\pi}^{\pi}W_T(\lambda-\mu)\mathbf{f}_j(u,\mu)d\mu - \mathbf{f}_j(u,\lambda)\right)\right\}d\lambda du\\
&\quad + \sqrt{T}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\lambda)\int_{-\pi}^{\pi}\left(\mathbf{I}_N(u,\mu)-\mathbf{f}_j(u,\mu)\right)W_T(\lambda-\mu)d\mu\right\}d\lambda du\\
&\qquad - \sqrt{T}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\mu)\left(\mathbf{I}_N(u,\mu)-\mathbf{f}_j(u,\mu)\right)\right\}d\mu du\\
&= \sqrt{T}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{Q}_{j,k}(u,\lambda)\left(\int_{-\pi}^{\pi}W_T(\lambda-\mu)\mathbf{f}_j(u,\mu)d\mu - \mathbf{f}_j(u,\lambda)\right)\right\}d\lambda du
\end{aligned}$$

$$+\sqrt{T}\int_0^1\int_{-\pi}^{\pi} tr\left\{\mathbf{D}(u,\mu)\left(\mathbf{I}_N(u,\mu)-\mathbf{f}_j(u,\mu)\right)\right\}d\mu du$$

$$= L_T^{(1)} + L_T^{(2)} \quad (say), \tag{96}$$

where

$$\mathbf{D}(u,\mu) = \int_{-\infty}^{\infty}\left\{\mathbf{Q}_{j,k}\left(u,\mu+\frac{x}{M}\right)-\mathbf{Q}_{j,k}(u,\mu)\right\}W(x)dx. \tag{97}$$

By the dominated convergence theorem,

$$\lim_{T\to\infty}\|\mathbf{D}(u,\mu)\| = 0 \quad a.e.\ (\mu\in[-\pi,\pi]), \tag{98}$$

therefore, from Lemma 2, $Var\{L_T^{(2)}\} = o(1)$, which implies $L_T^{(2)} = o_p(1)$. On the other hand, by Assumption 3, we have

$$\int_{-\pi}^{\pi}\mathbf{f}(u,\mu)W_T(\lambda-\mu)d\mu - \mathbf{f}(u,\lambda) = O(M^{-2}), \tag{99}$$

hence $L_T^{(1)} = O\left(\frac{\sqrt{T}}{M^2}\right) = o(1)$. □

Proof of Theorem 2. From Assumptions 6, 7 and

$$\mathbf{f}_j(u,\lambda)\mathbf{f}_k(u,\lambda)^{-1} = \mathbf{E}_m + \frac{(-1)^j}{\sqrt{T}}\mathbf{\Delta}(u,\lambda) + O(T^{-1}), \tag{100}$$

it is seen that

$$D_H(\mathbf{f}_j;\mathbf{f}_k) = \frac{c}{8\pi T}\int_0^1\int_{-\pi}^{\pi} tr\{\mathbf{\Delta}(u,\lambda)\}^2 d\lambda du + o(T^{-1}) \tag{101}$$

and

$$\begin{aligned} V_H^2(j,k) &= \frac{c^2}{4\pi T}\int_0^1\int_{-\pi}^{\pi} tr\{\mathbf{\Delta}(u,\lambda)\}^2 d\lambda du \\ &+\frac{c^2}{64\pi^4 T}\sum_{a_1,a_2,a_3,a_4=1}^{m}\kappa_{a_1,a_2,a_3,a_4}\gamma_{a_1,a_2}\gamma_{a_3,a_4} + o(T^{-1}), \end{aligned} \tag{102}$$

where $\gamma_{a,b}$ is the (a,b)-th element of the $m\times m$ matrix

$$\begin{aligned} \mathbf{\Gamma}_H = \int_0^1\int_{-\pi}^{\pi} & \mathbf{A}^{(1)}(u,\lambda)\{\mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^*\}^{-1} \\ & \{\mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(2)}(u,\lambda)^* + \mathbf{A}^{(2)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^*\} \\ & \{\mathbf{A}^{(1)}(u,\lambda)\mathbf{\Omega}\mathbf{A}^{(1)}(u,\lambda)^*\}^{-1}\mathbf{A}^{(1)}(u,\lambda)d\lambda du. \end{aligned} \tag{103}$$

If $\mathbf{\Gamma}_H = \mathbf{0}$, substituting them into (41), then the asymptotic misclassification probabilities are given by

$$\begin{aligned} &\lim_{T\to\infty}P_{D_H}(2|1) = \lim_{T\to\infty}P_{D_H}(1|2) \\ &= \Phi\left(-\frac{1}{2}\sqrt{\frac{1}{4\pi}\int_0^1\int_{-\pi}^{\pi} tr\{\mathbf{\Delta}(u,\lambda)\}^2 d\lambda du}\right). \end{aligned} \tag{104}$$

Since

$$\mathbf{\Gamma}_H = \int_0^1 \int_{-\pi}^{\pi} \{\mathbf{A}^{(2)}(u,\lambda)^* \mathbf{A}^{(1)}(u,\lambda)^{*^{-1}} \mathbf{\Omega}^{-1} + \mathbf{\Omega}^{-1}\mathbf{A}^{(1)}(u,\lambda)^{-1}\mathbf{A}^{(2)}(u,\lambda)\} d\lambda du$$

$$= 2\pi \int_0^1 \left\{ \mathbf{a}_0^{(2)}(u)' \mathbf{a}_0^{(1)}(u)^{-1'} \mathbf{\Omega}^{-1} + \mathbf{\Omega}^{-1}\mathbf{a}_0^{(1)}(u)^{-1}\mathbf{a}_0^{(2)}(u) \right\} du, \tag{105}$$

$$vec\ \mathbf{\Gamma}_H = 2\pi \int_0^1 vec\ \left\{ \{\mathbf{K}_m + \mathbf{E}_m\} \mathbf{\Omega}^{-1}\mathbf{a}_0^{(1)}(u)^{-1}\mathbf{a}_0^{(2)}(u) \right\} du$$

$$= 4\pi \int_0^1 \{\mathbf{E}_m \otimes \mathbf{\Omega}^{-1}\mathbf{a}_0^{(1)}(u)^{-1}\} vec\ \mathbf{a}_0^{(2)}(u) du, \tag{106}$$

which implies that $\mathbf{\Gamma}_H = \mathbf{0}$ is equivalent to $\mathbf{a}_0^{(2)}(u) \equiv \mathbf{0}$. □

Proof of Theorem 3. Under Π_1, it is seen that

$$\begin{aligned}
&\Lambda_T(p_1,p_2) - F_T(p_1,p_2) \\
&= \sum_{t=1}^T \left\{ \log p\left(\varepsilon_t\right) - \log p \left(\frac{\sum_{k=0}^{t-1} b^\circ_{\theta_T,t,T}(k) X_{t-k,T} + \sum_{r=0}^\infty c^\circ_{\theta_T,t,T}(r)\varepsilon_{-r}}{a^\circ_{\theta_T,t,T}(0)} \right) \right. \\
&\qquad \left. - \log p \left(\frac{\sum_{k=0}^{t-1} b^\circ_{\theta,t,T}(k) X_{t-k,T}}{a^\circ_{\theta,t,T}(0)} \right) + \log p \left(\frac{\sum_{k=0}^{t-1} b^\circ_{\theta_T,t,T}(k) X_{t-k,T}}{a^\circ_{\theta_T,t,T}(0)} \right) \right\} \\
&= \sum_{t=1}^T \left\{ q_{t,T}\phi\left(\varepsilon_t\right) + \frac{q_{t,T}^2}{2} D\phi\left(\varepsilon_t\right) - r_{t,T}\phi \left(\frac{\sum_{k=0}^{t-1} b^\circ_{\theta,t,T}(k) X_{t-k,T}}{a^\circ_{\theta,t,T}(0)} \right) \right. \\
&\qquad \left. - \frac{r_{t,T}^2}{2} D\phi \left(\frac{\sum_{k=0}^{t-1} b^\circ_{\theta,t,T}(k) X_{t-k,T}}{a^\circ_{\theta,t,T}(0)} \right) \right\} + O_P\left(T^{-1/2}\right) \\
&= \sum_{t=1}^T \left\{ q_{t,T}\phi\left(\varepsilon_t\right) + \frac{q_{t,T}^2}{2} D\phi\left(\varepsilon_t\right) - r_{t,T}\phi\left(\varepsilon_t\right) - \frac{r_{t,T}^2}{2} D\phi\left(\varepsilon_t\right) \right\} \\
&\quad + O_P\left(T^{-1/2}\right) + o_P\left(1\right),
\end{aligned} \tag{107}$$

where

$$\begin{aligned}
q_{t,T} &= r_{t,T} + \frac{\sum_{r=0}^\infty c^\circ_{\theta,t,T}(r)\varepsilon_{-r}}{a^\circ_{\theta,t,T}(0)} - \frac{\sum_{r=0}^\infty c^\circ_{\theta_T,t,T}(r)\varepsilon_{-r}}{a^\circ_{\theta_T,t,T}(0)} \\
&= r_{t,T} + \frac{h'}{\sqrt{T}} \sum_{r=0}^\infty \left\{ \frac{\nabla c^\circ_{\theta^*,t,T}(r)}{a^\circ_{\theta,t,T}(0)} + \frac{c^\circ_{\theta_T,t,T}(r) \nabla a^\circ_{\theta^{**},t,T}(0)}{a^\circ_{\theta,t,T}(0) a^\circ_{\theta_T,t,T}(0)} \right\} \varepsilon_{-r} \\
&= r_{t,T} + O\left(T^{-1/2} t^{-1}\right)
\end{aligned} \tag{108}$$

and

$$\begin{aligned}
r_{t,T} &= \frac{\sum_{k=0}^{t-1} b^\circ_{\theta,t,T}(k) X_{t-k,T}}{a^\circ_{\theta,t,T}(0)} - \frac{\sum_{k=0}^{t-1} b^\circ_{\theta_T,t,T}(k) X_{t-k,T}}{a^\circ_{\theta_T,t,T}(0)} \\
&= \frac{h'}{\sqrt{T}} \left\{ \sum_{k=1}^{t-1} \frac{\nabla b^\circ_{\theta^{***},t,T}(k)}{a^\circ_{\theta,t,T}(0)} X_{t-k,T} + \sum_{k=0}^{t-1} \frac{b^\circ_{\theta_T,t,T}(k) \nabla a^\circ_{\theta^{**},t,T}(0)}{a^\circ_{\theta,t,T}(0) a^\circ_{\theta_T,t,T}(0)} X_{t-k,T} \right\}.
\end{aligned} \tag{109}$$

Here θ^*, θ^{**} and θ^{***} are points on the segment between θ and $\theta_T = \theta + h/\sqrt{T}$.
From (107), (108) and (109) we can see that $\Lambda_T(p_1, p_2) - F_T(p_1, p_2) = o_P(1)$ under Π_1. Similarly, we have $\Lambda_T(p_1, p_2) - F_T(p_1, p_2) = o_P(1)$ under Π_2. Therefore, $F_T(p_1, p_2)$ has the same limit distribution of $\Lambda_T(p_1, p_2)$ under both Π_1 and Π_2. $\square$

Acknowledegments The author would like to express his sincere thanks to Professor Masanobu Taniguchi for his encouragement and guidance. He is also grateful to a referee for his/her helpful comments and advices.

References

[1] Anderson, T. W,, 1984. *An introduction to Multivariate Statistical Analysis.* New York: Wiley.

[2] Dahlhaus, R., 1996a. On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Processes and Their Applications* **62** 139-168.

[3] Dahlhaus, R., 1996b. Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparametr. Statist.* **6** 171-191.

[4] Dahlhaus, R., 1996c. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. *Athens Conference on Applied Probability and Time Series 2. Lecture Notes in Statist.* **115** 145-159. New York: Springer.

[5] Dahlhaus, R., 1997. Fitting time series models to nonstationary process. *Ann. Statist.* **25** 1-37.

[6] Dahlhaus, R., 2000. A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762-1794.

[7] Hirukawa, J. and Taniguchi, M. 2004. LAN Theorem for Non-Gaussian Locally Stationary Processes and Its Applications. Waseda University Time Series Discussion Paper 2.

[8] Kakizawa, Y., 1997. Discriminant analysis for non-Gaussian vector stationary processes. *J. Japan Statist. Soc.* **27.** 19-35.

[9] Kakizawa, Y., Shumway, R. H. and Taniguchi, M., 1998. Discrimination and Clustering for Multivariate Time Series. *J. Amer. Statist. Assoc.* **93.** 328-340.

[10] Magunus, J. R. and Neudecker, H., 1988. *Matrix Differential Calculus With Applications in Statistics and Econometrics.* Chichester, U. K.: Wiley.

[11] Parzen, E., 1990. Time Series, Statistics and Information. *New directions in time series analysis, Part I. IMA Vol. Math. Appl.* **45.** 265-286. New York: Springer.

[12] Sakiyama, K., 2002. Some statistical applications for locally stationary processes. *Sci. Math. Jpn.* **56.** 231-250

[13] Sakiyama, K. and Taniguchi, M., 2004. Discriminant Analysis for Locally Stationary Processes. To appear in *J. Multivariate Anal.*

[14] Shumway, R. H., 1982. Discriminant analysis for time series. *Handbook of Statistics. Vol.2.* 1-46. Amsterdam: North-Holland.

[15] Shumway, R. H., 2003. Time-frequency clustering and discriminant analysis. *Statist. Probab. Lett.* **63.** 307-314.

[16] Taniguchi, M. and Kakizawa, Y., 2000. *Asymptotic Theory of Statistical Inference for Time Series.* New York: Springer.

[17] Taniguchi, M. Puri, M. L. and Kondo, M., 1996. Nonparametric Approach for Non-Gaussian Vector Stationary Processes. *J. Multivariate Anal.* **56.** 259-283.

[18] Zhang, G. and Taniguchi, M., 1994. Discriminant Analysis for Stationary Vector Series. *J. Time Series Anal.* **15.** 117-126.

[19] Zhang, G. and Taniguchi, M., 1995. Nonparametric Approach for Discriminant Analysis in Time Series. *J. Nonparametr. Statist.* **5.** 91-101.

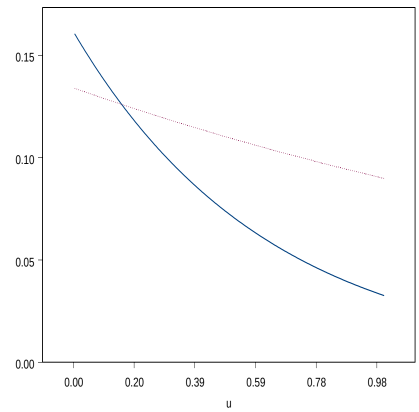

Figure 1: The time varying coefficient functions $b_\theta^{(1)}(u)$ (real line) and $b_\theta^{(2)}(u)$ (dotted line).

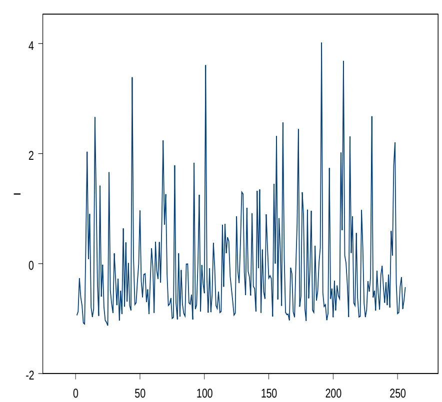

Figure 2: The observation $\{X_{t,T}\}$ of Example 1.

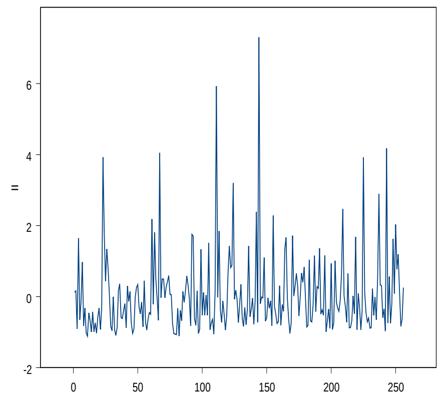

Figure 3: The observation $\{X_{t,T}\}$ of Example 2.

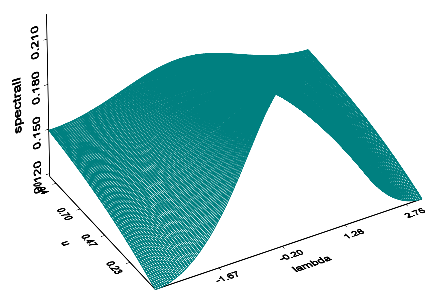

Figure 4: The time varying spectral density function $f(u, \lambda)$ for Example 1.

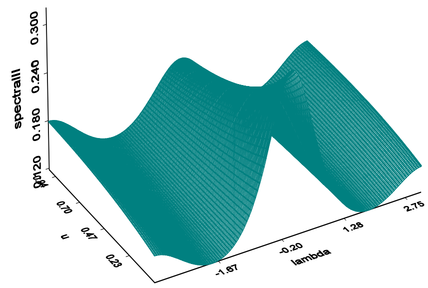

Figure 5: The time varying spectral density function $f(u, \lambda)$ for Example 2.

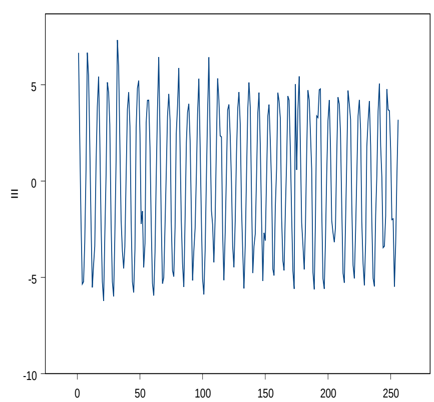

Figure 6: The observation $\{Y_{t,T}\}$ which is contaminated by a sharp peak.

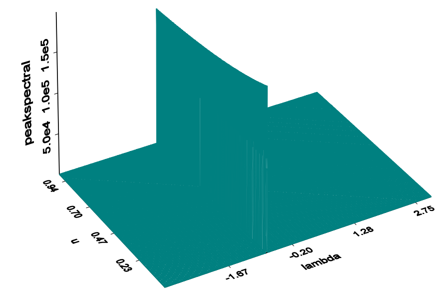

Figure 7: The time varying spectral density $\bar{f}_1(u, \lambda)$ which is contaminated by a sharp peak.

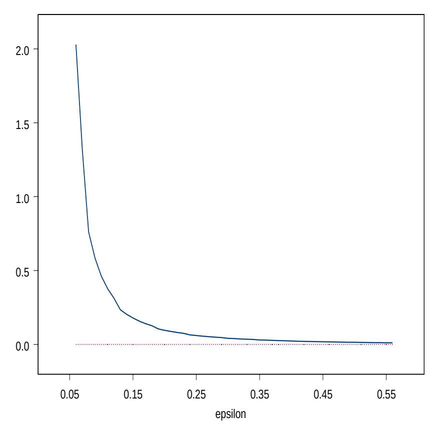

Figure 8: The pair of graphs $K(\bar{f}_1, f_1, f_2)$ (real line) and $K(f_1, f_2)$ (dotted line).

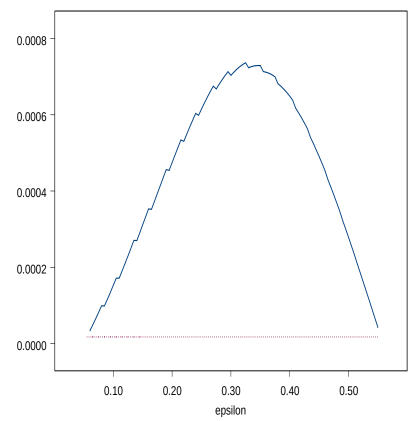

Figure 9: The pair of graphs $B_{1/2}(\bar{f}_1, f_1, f_2)$ (real line) and $B_{1/2}(\bar{f}_1, f_1, f_2)$ (dotted line).

*Department of Mathematical Sciences School of Science & Engineering
Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo,169-8555 Japan
E-mail: hirukawa@ruri.waseda.jp